

*Marcin Kuraś*Université Nicolas Copernic
de ToruńLES MÉMOIRES DE TRADUC-
TION : OUTIL NÉCESSAIRE
OU SOURCE DE PROBLÈMES
POUR LES TRADUCTEURS ?

1. INTRODUCTION

Les mémoires de traduction, comprises comme des bases de données contenant des segments des textes de la langue source et leurs segments correspondants de la langue cible ont été créées comme un des composants des systèmes de la traduction automatique. Leurs principes de base ont été publiés dans le fameux rapport d'ALPAC (Automatic Language Processing Advisory Committee), daté de 1966¹, dont la conclusion générale était l'impossibilité de construire un système de la traduction automatique produisant des résultats efficaces sans intervention humaine. Les auteurs du rapport ont constaté que la traduction faite par un ordinateur nécessitait toujours la correction par le traducteur humain pour que sa qualité ait été acceptable. Ce processus exigeait beaucoup plus d'effort (voire l'argent) qu'une traduction traditionnelle. En résultat, le développement et les recherches sur la traduction automatique, financés par l'état, ont été arrêtés pour presque vingt ans.

Cependant, les recherches pour faciliter et améliorer le processus de la traduction lui-même continuaient dans le domaine de la traduction assistée par ordinateur (TAO). Ces actions ont été supportées par des organisations qui devaient traduire un grand nombre de documents en peu de temps et en bonne qualité. À l'époque de la publication du rapport d'ALPAC c'était le Département de la Défense des États-Unis qui a développé les glossaires textuels pour l'Agence de Traduction des Forces Armées à Mannheim et la Communauté Européenne du Charbon et de l'Acier qui, en 1975, a créé l'EURODICAUTOM : la base de données de terminologie étant définie aux contextes des phrases. Cette version initiale a été remplacée en 2007 par Inter-Active Terminology for Europe (la base de données terminologique multilingue de l'Union Européenne) qui à présent contient plus de 8 millions de termes, environ 540 milles d'abréviations et 130 milles d'expressions, couvrant 23 langues officielles de l'UE². Cette base peut être gratuitement consultée en ligne sous l'adresse : <http://iate.europa.eu>. Actuellement, les organismes de l'Union Européenne contribuent

¹ Sur son histoire trent ans après, voir Hutchins (1996).

² http://iate.europa.eu/iatediff/brochure/IATEbrochure_FR.pdf (4.10.2010).

aussi au développement des systèmes de la traduction automatique (TA) qui, à l'heure actuelle, atteignent d'ailleurs des résultats de plus en plus satisfaisants.

2. STRUCTURE ET FONCTIONNEMENT DES SYSTÈMES DE TAO

L'amélioration des relations « prix / qualité / vitesse » dans le domaine de traduction des textes, c'est le motif primordial du développement des mémoires de traduction qui ont bientôt évaluées de simple glossaire vers les systèmes complexes de TAO. Il y en a plusieurs sur le marché actuel, mais ils visent tous, entre autres, à fournir au traducteur la possibilité d'utiliser automatiquement pendant la traduction une base de données des segments des textes traduits antérieurement. Cela semble être une aide excellente pour des traducteurs. Simple de principe et efficace. Pourtant, comme chaque technologie, celle-là possède aussi ses limitations. Tout d'abord, au but d'analyser l'utilité des systèmes de TAO, il faut se rendre compte de leur fonctionnement et structure. Chaque système de TAO est composé de plusieurs modules qui offrent des fonctions différentes.

2.1. L'ENVIRONNEMENT DE TRAVAIL

Premièrement, au niveau de l'interface, c'est le module du traitement de texte qui permet la création, l'édition et la correction des documents. On peut distinguer ici trois catégories d'outils TAO : ceux qui s'intègrent avec le logiciel de traitement de texte le plus populaire, c'est-à-dire le MS Word™ (et fonctionnent comme des macros) ; ceux qui offrent leurs propres modules de traitement (et fonctionnent comme un seul logiciel intégrant toutes les fonctions) et ceux qui permettent d'éditer les textes à partir d'un navigateur Web.

■ La I^{ère} catégorie est la solution la plus acceptable surtout par les traducteurs qui débutent avec les systèmes TAO, puisque cela leur offre le confort de travailler en utilisant le milieu qu'ils connaissent bien. Néanmoins, pour profiter de l'avantage de mémoire de traduction, il faut d'habitude travailler avec plusieurs logiciels ouverts simultanément (la fenêtre de Word™, l'interface du mémoire de traduction, la fenêtre avec le glossaire terminologique, etc.). On peut qualifier ici de tels systèmes comme Metatexis™, Wordfast™ (jusqu'à la version 5), Trados™ (jusqu'à la version 2007), WebWordsSystem™, etc.

■ La II^e catégorie (et la troisième aussi) exige un peu plus d'effort pour maîtriser la manière différente de travailler avec en texte à traduire. Pourtant, dès que cela est faite, le travail devient plus facile et efficace. Particulièrement, parce que le traducteur ne doit se préoccuper ni du format de fichier source et cible (que ce soit DOC, PDF, XLS, PPT, etc.) ni de la mise en page qu'il faut conserver. Le logiciel fait tout pour le traducteur : il importe les fichiers source dans le système, présente le contenu d'un document sous une forme facile à traduire (en omettant les éléments non-textuels) et finalement veille à ce que le fichier cible soit exporté avec la même mise en page que le document original. La règle générale de fonctionnement du module d'édition est

pareille pour tous les systèmes intégrés. D'habitude la fenêtre du logiciel est divisée en plusieurs parties : la partie d'édition/traduction (ou on voit les segments sources et on les traduit), la partie de mémoire de traduction (où les résultats des recherches des traductions d'un segment donné sont affichés) et la partie de terminologie (où le glossaire terminologique est affiché). Optionnellement, on peut distinguer aussi la partie d'aperçu (où on peut voir la structure et la mise en page du document original) et la partie de navigation (dans le fichier ou dans le logiciel). Comme les exemples des systèmes intégrés on peut citer MemoQ™, SDL Trados Studio™, Across™, Transit NXT™ ou Déjà Vu™.

■ La III^e catégorie de systèmes permet de travailler en ligne, tout simplement avec un navigateur Web. Dans ce cas, le traducteur ne doit rien installer et il peut travailler n'importe où, même sans avoir son propre ordinateur. D'autre part, ces systèmes n'offrent pas une telle flexibilité que ceux qui sont installés localement. Ici, il faut mentionner surtout :

- le fameux « Google Kit du traducteur » (<http://translate.google.com/toolkit>),
- Wordfast Enywhere™ (<http://anywhere.wordfast.net/>),
- Ontram™ (<http://www.andrae-ag.com/>).

2.2. SEGMENTATION

Ensuite, dès que le fichier est importé au système, un autre module, celui de segmentation devient actif. Son rôle est de partager le texte source en unités (TU – Translation Unit) et il le fait d'après les règles de segmentation qui précisent comment traiter le texte, distinguer les phrases et paragraphes, reconnaître les signes de ponctuation, nombres, etc. Malheureusement, ces règles sont différentes pour chaque système et l'utilisateur d'habitude ne peut rien faire pour les modifier. Il y a certainement des systèmes qui permettent de changer les règles de segmentation comme par exemple WordFast™ ou Sisulizer™, mais leur création ou leur modification est difficile. Par conséquent, le même texte peut être segmenté différemment par plusieurs logiciels. Parfois, cela pose un grand problème pour un traducteur qui est forcé à traduire des unités illogiques (phrases coupées au milieu). En conséquence, les paires des segments sources et traduits ne sont pas utiles dans le procès de la traduction. C'est par exemple le cas de « Google Kit du traducteur ». Ce système en ligne, d'ailleurs très utile, n'est pas capable de bien segmenter le texte contenant les listes ordonnées et non-ordonnées. En outre, dès que le fichier est importé au système, l'utilisateur ne peut rien modifier dans le texte source. Un autre problème de la segmentation résulte de la mise en page incorrecte du document source. Si le fichier n'est pas proprement préparé, l'algorithme de segmentation de Google n'est pas capable de bien partager le texte en unités.

Par contre, en 2004, une association LISA (Localization Industry Standards Association) a adopté la première version du standard SRX (Segmentation Rules eXchange) qui avait pour but d'unifier les règles de segmentation et de permettre le transfert de ces règles entre les systèmes différents. En 2008, la seconde version de SRX a apparu³. La question de standardisation sera abordée d'une façon plus détaillée dans la suite de ce travail.

³ <http://www.lisa.org/Segmentation-Rules-eXchange-SRX.40.0.html> (10.10.2010).

2.3. MÉMOIRE DE TRADUCTION

Quand le texte est correctement segmenté, le traducteur peut directement commencer son travail. En ce moment-là, il peut profiter de l'avantage du système de TAO sur un simple logiciel du traitement de texte puisque il gagne l'accès à des bases de données des unités lexicales déjà traduites en langue cible. On peut distinguer trois types de telles bases (mémoires de traduction) : la base locale (privée), créée auparavant par le traducteur et modifiée en cours de la traduction quand on y ajoute les unités nouvelles récemment traduites; la base en ligne, appelée un VLTM (Very Large Translation Memory / une Très Large Mémoire de Traduction) et le glossaire de la terminologie (obligatoire ou interdite). Le module de mémoire de traduction intègre ces trois bases de données et il suggère au traducteur la solution pour traduire un segment donné à condition que ce segment identique du texte source ou un segment similaire est retrouvé. Dans le cas de segments qui ne sont pas identiques, le logiciel calcule le taux de similarité entre le segment source original et le segment du mémoire de traduction en présentant le pourcentage approprié. Le traducteur peut accepter la solution (ou les solutions) proposée ou la rejeter. Il est également possible de définir le seuil de similitude, c'est-à-dire le pourcentage en dessous duquel la traduction est considérée comme incorrecte et, par conséquent, n'est pas affichée. La méthode de calcul est secrète et diffère entre les systèmes de TAO, alors les taux de similarité ne sont pas comparables. Dans la plupart des cas, l'algorithme est uniquement mathématique (statistique) et l'apport de la linguistique comme l'analyse syntaxique et sémantique est ignoré. Ce technique, appelé la correspondance floue (fuzzy matching⁴), est à l'heure actuelle le plus implémenté.

Par contre, récemment une approche plus intelligente est apparue. C'est l'exemple de MetaMorphoTM, développé par Gábor Hodász et Gábor Pohl à l'Université Catholique de Péter Pázmány à Budapest, en coopération avec la société MorphoLogic. Ce système utilise l'analyse linguistique des segments en langue source et en langue cible enregistrés dans la mémoire de traduction pour produire la suggestion de traduction d'une phrase donnée plus exacte que celle obtenue par des méthodes purement statistiques (Hodász, Pohl 2005). Cette analyse s'effectue à plusieurs niveaux : celui d'une forme du mot, celui d'un lemme du mot et celui d'une catégorie grammaticale du mot. De surcroît, les entrées dans la mémoire de traduction (les segments) sont transmises en phrases nominales et alignés ensuite.

2.4. ALIGNEMENT

Évidemment, pour utiliser la mémoire de traduction locale comme une aide pendant la traduction d'un texte nouveau, il faut la préparer au préalable. Si on a déjà travaillé avec le système de TAO pour traduire un autre texte du même domaine, une telle base de données a été créée automatiquement. En revanche, si on a effectué la traduction traditionnelle et on possède deux fichiers : document original en langue source et le

⁴ http://en.wikipedia.org/wiki/Fuzzy_matching (8.11.2010).

document traduit, au but de construire une mémoire de traduction il est nécessaire de réaliser une opération d'alignement. Pour cela il faut profiter d'un module d'alignement qui est la partie intégrante de tous les systèmes de TAO. Ce travail est malheureusement considéré comme pénible, puisqu'il exige beaucoup de temps à l'exécuter. Par contre, l'alignement des documents au but de créer une mémoire de traduction à utiliser pour une nouvelle traduction peut considérablement réduire le temps consacré au travail. C'est surtout le cas de la traduction d'une grande quantité des documents du même domaine, où il y a beaucoup de répétitions et la terminologie homogène doit être conservée, comme par exemple les textes techniques, juridiques, etc.

2.5. TRADUCTION AUTOMATIQUE

Un autre module optionnel qui est intégré dans presque tout les systèmes de TAO c'est celui de la traduction automatique. D'habitude c'est juste une interface entre la mémoire de traduction et le système de TA disponible localement ou en ligne. C'est par exemple le cas de MemoQTM ver. 4.5 ou SDL Trados Studio 2009TM qui utilisent le mécanisme Google TranslateTM. Autres, comme MultitransTM profitent du SystranTM. Le module de TA est parfois utile si un segment à traduire n'est pas retrouvé dans la mémoire de traduction (le taux de similitude est inférieur à un seuil établi par l'utilisateur). Les systèmes de TA modernes, (comme Google TranslateTM, SDL Language WeaverTM ou Bing TranslatorTM), basés sur l'approche statistique dans l'analyse des corpus multilingues, offrent souvent les traductions de qualité acceptable, au moins au niveau des phrases simples et des expressions.

À vrai dire, on observe actuellement le retour vers les systèmes de la traduction automatique dans la traduction assistée par ordinateur.

3. STANDARISATION – LISA

Au moment de création de premiers systèmes de TAO dans les années 80, chaque logiciel avait son propre solution pour des mémoires de traduction, les règles de segmentation et la gestion de terminologie. Par conséquent, les données n'étaient pas interchangeable parmi eux. Cette situation continuait jusqu'aux années 90, quand les premiers essais de standardisation ont apparus.

L'Association LISA (Localization Industry Standards Association) a été fondée en 1990 au but de permettre les entreprises et les personnes individuelles, s'occupant d'adaptation des produits divers aux marchés locaux et de créer une communauté d'échange d'expériences⁵. Dans la structure de LISA fonctionnent plusieurs comités et, en particulier, l'OSCAR (Open Standards Container/Content Allowing Reuse) : le comité de standardisation pour l'industrie de localisation et traduction⁶. Le comité OSCAR depuis son instauration a développé plusieurs standards pour la technologie de

⁵ <http://www.lisa.org/Why-LISA.111.0.html> (8.11.2010).

⁶ <http://www.lisa.org/OSCAR-LISA-s-Standards-Committee.79.0.html> (8.11.2010).

traduction et de localisation, notamment : le TMX (Translation Memory eXchange), qui sert à échanger les mémoires de traduction créées à partir des outils TAO ; le SRX (Segmentation Rules eXchange) utilisé pour la description des règles de segmentation d'un outil au but de les transmettre à un autre et le TBX (Term-Base eXchange), servant à exporter les glossaires terminologiques. Ces trois standards basent sur le langage XML⁷ (Extensible Markup Language – Langage Extensible de Balisage) et étaient très importants pour le développement des systèmes de TAO.

4. CONCLUSION

Le procès de traduction effectué par un être humain est, toutes proportions gardées, comparable à celui fait par un outil de mémoire de traduction. Excepté, bien sûr, l'apport créatif d'un homme, puisque les systèmes de TAO ne sont pas intelligents. Un homme tout d'abord lit le texte à traduire et essaye de le comprendre. Ensuite, il choisit les unités à traduire, il cherche la terminologie appropriée et il traduit le texte, en étant forcé à retenir dans sa mémoire les phrases et les structures lexicales déjà traduits qui se répètent. Par contre, un outil de TAO segmente le texte en unités à traduire, enregistre dans sa base de données les éléments déjà traduits, compare les unités nouvelles avec celles de la base et finalement suggère la traduction des unités similaires. Autrement dit, une mémoire de traduction automatise le procès de comparaison des unités du texte source avec la traduction dans le texte cible, en servant comme une extension de la mémoire d'un traducteur (d'où, d'ailleurs, vient le nom : mémoire de traduction).

Par conséquent, l'usage d'un système de TAO sera le plus profitable si un traducteur est obligé de traduire régulièrement de grandes quantités de textes. Surtout, au cas où ces textes concernent le même domaine et leur lexique est plus ou moins homogène, ce qui garantit le taux de répétitions assez haut. Dans ce cas-là, le travail consacré initialement à la préparation des mémoires de traduction (bases de données des segments ou les glossaires terminologiques) nous économisera beaucoup de temps au moment de la traduction.

Il faut mentionner ici la différence entre la traduction, définie comme le simple transfert du texte de la langue source vers la langue cible et la localisation, généralement définie comme la traduction de l'interface utilisateur. Cette tâche exige l'usage des systèmes de TAO destinés uniquement à la localisation, comme PassoloTM, Alchemy CatalystTM ou SDL InsightTM et n'est pas possible à exécuter en se servant des méthodes traditionnelles de la traduction.

Pour conclure, on peut dire que les outils de TAO sont nécessaires est très utile pour des traducteurs professionnels conscients de leur fonctionnalité. Pourtant, les traducteurs occasionnels ou ceux qui préfèrent le travail traditionnel y trouveront uniquement la source des problèmes.

⁷ <http://fr.wikipedia.org/wiki/XML> (9.11.2010).

BIBLIOGRAPHIE

- CHENON Christophe, 2005, *Vers une meilleure utilisabilité des mémoires de traduction, fondée sur un alignement sous-phrastique*, Thèse doctorale présentée à l'Université Joseph Fourier – Grenoble 1, disponible sur : <http://hal.inria.fr/docs/00/06/56/39/PDF/TheseXoFChenon-31oct05.pdf> (20.12.2011).
- CORDOVA Barbara, *La Mémoire de Traduction : Un Outil de Traduction Assistée Par Ordinateur*, disponible sur : <http://www.translationdirectory.com/article313.htm> (20.12.2011).
- HODÁSZ Gábor, POHL Gábor, 2005, *MetaMorpho TM: a linguistically enriched translation memory*, conférence présentée dans le cadre de l'atelier « Modern Approaches in Translation Technologies » (Borovets, Bulgarie), texte disponible sur : www.mt-archive.info?RANLP-2005-Hodasz.pdf (20.12.2011).
- HUTCHINS John, 1996, ALPAC: the (in)famous report, *MT News International 14*, Juin, 9–12, disponible sur : www.hutchinsweb.me.uk/MTNI-14-1996.pdf (20.12.2011).
- PRÓSZÉKY Gábor, 2010, *Some Thoughts on the Future of Machine Translation*, conférence dans le cadre du programme TM-Europe 2010 (Kraków 29.09.–1.10.2010).
- ZERFASS Angelika, 2010a, *Translation Tools Primer*, conférence dans le cadre du programme TM-Europe 2010 (Kraków 29.09.–1.10.2010).
- ZERFASS Angelika., 2010b, *Trends in Translation Technologies*, conférence dans le cadre du programme TM-Europe 2010 (Kraków 29.09.–1.10.2010).

Summary

Translation Meemories: a necessary tool or a source of problems for translators

The Translation Memories can be characterised as databases of text fragments in source language with their equivalents in target language that make it possible to effectuate the translation process in a more effective and automatic way. This is a tool that is a part of Computer Aided Translation (CAT) domain and at present seems to be an excellent solution for translators. However, there are also some disadvantages concerning the usage of those mechanisms. This short analysis tends toward determining who can use such software and in what case will it be most effective in creating various types of translation.

Streszczenie

Pamięci tłumaczeniowe: konieczne narzędzie czy źródło problemów dla tłumaczy?

Pamięć tłumaczeniowa (Translation Memory) może być scharakteryzowana jako baza danych zawierająca fragmenty tekstów w języku źródłowym wraz z odpowiadającymi im fragmentami w języku docelowym, co z założenia ma umożliwić bardziej efektywny i zautomatyzowany proces tłumaczenia tekstów. Narzędzia te są częścią dziedziny określanej jako tłumaczenie wspomagane maszynowo (komputerowo) i są obecnie uważane za doskonałe rozwiązanie dla tłumaczy. Niniejsza krótka analiza ma na celu wykazanie, dla kogo jest przeznaczone tego typu oprogramowanie i jakie systemy dostępne na rynku są najbardziej efektywne przy wykonywaniu różnych typów tłumaczeń.